

EXHIBIT 5

**UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK**

AUTHORS GUILD, et al.

Plaintiffs,

v.

OPEN AI INC., et al.,

Defendants.

ECF CASE

No. 1:23cv-08292-SHS-OTW

No. 1:23-cv-10211-SHS-OTW

JONATHAN ALTER, et al.

Plaintiffs,

v.

OPEN AI INC., et al.,

Defendants

**DEFENDANT MICROSOFT
CORPORATION’S RESPONSES AND
OBJECTIONS TO PLAINTIFFS’
SECOND SET OF REQUESTS FOR
PRODUCTION OF DOCUMENTS**

**DEFENDANT MICROSOFT CORPORATION’S RESPONSES AND OBJECTIONS TO
PLAINTIFFS’ SECOND SET OF REQUESTS FOR PRODUCTION OF DOCUMENTS**

Pursuant to Federal Rules of Civil Procedure 26 and 34, Defendant Microsoft Corporation (“Microsoft”) responds to Plaintiffs’ Second Set of Requests for Production of Documents (“Requests”) as follows:

GENERAL OBJECTIONS

Microsoft’s responses are subject to the following objections to the “Definitions” and “Instructions” provided with the Requests, each of which is hereby incorporated by reference into Microsoft’s responses below. If Microsoft objects to a definition of a term and that term is used in the definition of a subsequent term, Microsoft’s objections to the term used in the subsequent definition are incorporated by reference therein.

1. Microsoft objects to the definition of the term “Communication(s)” to the extent it is broader than or inconsistent with the definition provided in Local Civil Rule 26.3, which provides for uniform definitions in discovery requests. For the purposes of responding to the Requests, Microsoft will construe the term “Communication(s)” to mean “transmittal of information (in the form of facts, ideas, inquiries, or otherwise)” as defined in Local Civil Rule 26.3.

2. Microsoft objects to the definition of the term “Document(s)” to the extent it is broader than or inconsistent with the definition provided in Local Civil Rule 26.3, which provides for uniform definitions in discovery requests. For the purposes of responding to the Requests, Microsoft will construe the term “Document(s)” to be synonymous in meaning and equal in scope to the usage of the term “documents or electronically stored information” in Federal Rule of Civil Procedure 34(a)(1)(A), as defined in Local Civil Rule 26.3.

3. Microsoft objects to the definition of the terms “You,” “Your,” and “Microsoft” on the grounds that they are overly broad, unduly burdensome, not proportional to the needs of the case, seek information that is privileged or immune from discovery, and include persons or entities over which Microsoft has no control and, therefore, Microsoft does not have possession, custody, or control of information of any such persons or entities. For the purposes of responding to the Requests, Microsoft will interpret each of these terms to mean Microsoft Corporation.

4. Microsoft objects to the definition of “OpenAI” on the grounds that it is overly broad, unduly burdensome, and disproportionate to the needs of the case to the extent it includes entities other than the OpenAI entities listed as Defendants and persons other than the OpenAI Defendants’ present and former officers, directors, and employees acting in that capacity. For purposes of responding to the Requests, Microsoft will construe the term “OpenAI” to only include

the OpenAI Defendants and their present and former officers, directors, and employees, acting in that capacity.

5. Microsoft objects to the definition of the terms “Large Language Model,” “LLM,” “AI Model(s),” “Generative AI system(s),” and “model(s),” on the grounds that it is overly broad, unduly burdensome, and not proportional to the needs of the case to the extent that it includes generative artificial intelligence models that are not identified in the First Consolidated Class Action Complaint filed at Fiction Docket, No. 1:23-cv-08292-SHS, ECF No. 69 (the “Consolidated Complaint”) and that have no relationship to Microsoft. For purposes of responding to the Requests, Microsoft will interpret the terms “Large Language Model,” “LLM,” “AI Model(s),” “Generative AI system(s),” and “model(s),” to mean the specific large language models and generative artificial intelligence models identified in the Consolidated Complaint: GPT-3, GPT-3.5, GPT-4, and GPT-4 Turbo (“Relevant Generative AI Models”). Microsoft objects to the definition of the term “API Product(s)” on the grounds that it is overly broad, unduly burdensome, and not proportional to the needs of the case because they bear no relationship to the infringement allegations in the Consolidated Complaint, which are focused on the training of the Relevant Generative AI Models.

6. Microsoft objects to the definition of the term “Person(s)” to the extent it is broader than or inconsistent with the definition provided in Local Civil Rule 26.3, which provides for uniform definitions in discovery requests. For the purposes of responding to the Requests, Microsoft will construe the term “Person(s)” to mean “any natural person or any legal entity, including, without limit, any business or governmental entity or association” as defined in Local Civil Rule 26.3.

7. Microsoft objects to the definition of the terms “Fine Tune(d),” “Fine Tuning,” “Pre-Train(ed),” “Pre-Training,” “Train[ed],” and “Training” on the ground that it is vague and ambiguous in that it states that these terms “have the same meaning as the term discussed in Your website materials and statements,” and then allegedly quotes portions of OpenAI’s website and not Microsoft’s website.

8. Microsoft objects to the definition of “ChatGPT” on the grounds that it is overly broad, unduly burdensome, and not proportional to the needs of the case to the extent that it includes any consumer-facing applications developed by OpenAI that are not specifically identified in the Consolidated Complaint and to the extent that it incorporates the definition of “LLM.” Microsoft will interpret this term to mean the specific consumer-facing applications developed by OpenAI, and in which the underlying LLM(s) includes those set forth above, identified in the Consolidated Complaint: ChatGPT, ChatGPT Plus, and ChatGPT Enterprise.

9. Microsoft objects to the definition of the term “Concern” to the extent it is broader than or inconsistent with the definition provided in Local Civil Rule 26.3, which provides for uniform definitions in discovery requests. For the purposes of responding to the Requests, Microsoft will construe the term “Concern” to mean “relat[e] to, refer[] to, describ[e], evidenc[e], or constitut[e]” as defined in Local Civil Rule 26.3.

10. Microsoft objects to the definition of the term “Training Material” on the grounds that it is overly broad, unduly burdensome, and not proportional to the needs of the case to the extent that it includes any data other than text data that may have been used to train an LLM because the allegations in the Consolidated Complaint relate solely to text-based generative artificial intelligence models. For purposes of responding to the Requests, Microsoft will construe the term “Training Material” to be limited to text data.

11. Microsoft objects to the relevant time period of January 1, 2015, to present on the grounds that is overly broad, unduly burdensome, and not proportional to the needs of the case to the extent that it includes time frames that are unrelated to the allegations in the Consolidated Complaint. For purposes of responding to the Requests, Microsoft will conduct a reasonable search for documents starting January 1, 2018.

12. Microsoft objects to each of the Requests as vague and ambiguous because many of the Requests imply, suggest and/or assume that Microsoft trained or otherwise developed OpenAI's models. Microsoft did not train the Relevant Generative AI Models, does not possess the training datasets OpenAI used to train the Relevant Generative AI Models, and is unaware of the contents of the training datasets used by OpenAI to train the Relevant Generative AI Models. Microsoft will not search for or produce documents in response to Requests that imply, suggest and/or assume that Microsoft trained or otherwise developed OpenAI's models.

13. Microsoft objects to each Request to the extent it seeks information or documents that are subject to the attorney-client privilege, work product doctrine, joint defense privilege, common interest protection, or any other applicable privilege, immunity, or protection provided by law, or that are trial preparation materials (collectively "Privileged Materials"). Privileged Materials will not be produced.

14. Microsoft objects to all "Definitions and "Instructions" as overly broad, unduly burdensome, and not proportional to the needs of the case because they seek to impose upon Microsoft burdens in addition to or inconsistent with those set forth in the Federal Rules of Civil Procedure; the Local Rules of this Court, especially Local Civil Rule 26.3, which provides for uniform definitions in discovery requests; and any other applicable rules or laws. Microsoft will interpret these Definitions and Instructions as not requiring Microsoft to perform unreasonable

searches, not requiring cumulative or duplicative discovery, and not requiring Microsoft to exceed the obligations set forth in the applicable rules and laws.

15. Documents will be produced on a rolling basis in accordance with the schedule governing these Actions, as ordered by the Court.

RESPONSES TO REQUESTS FOR PRODUCTION

REQUEST FOR PRODUCTION NO. 41:

Documents sufficient to identify all datasets containing commercial works of fiction and/or nonfiction, including Class Works, that **You** or **OpenAI** accessed, downloaded or copied to train **Large Language Models**.

RESPONSE TO REQUEST FOR PRODUCTION NO. 41:

Microsoft incorporates by reference and reasserts its General Objections to Plaintiffs' "Definitions" set forth above. Microsoft objects to the term "copied" to the extent it assumes or implies that Microsoft or OpenAI created a "copy" as defined by the Copyright Act. Microsoft further objects to this Request to the extent it assumes that Microsoft trained or otherwise developed OpenAI's models or applications identified in the Consolidated Complaint. Microsoft did not train the Relevant Generative AI Models, does not possess the training datasets OpenAI used to train the Relevant Generative AI Models, and is unaware of the contents of the training datasets used by OpenAI to train the Relevant Generative AI Models. Microsoft further objects to this Request as irrelevant and/or vague, ambiguous, overly broad, and not proportional to the needs of the case because it requests information concerning all of Microsoft's and OpenAI's training datasets, pre-training datasets, and/or fine-tuning datasets, regardless of whether they have been identified in the Consolidated Complaint and/or bear any relation to the LLMs at issue in these Actions.

Subject to these objections, Microsoft responds as follows: Microsoft objects to this Request because it implies, suggests, and/or assumes that Microsoft trained or otherwise developed OpenAI's models. Microsoft has already made a thorough investigation (including custodial and non-custodial material) into the issue of whether it had any role in OpenAI's pretraining of its LLMs. As such, Microsoft is relying upon this objection for the purpose of limiting the scope of its investigation on this particular topic to the reasonable search it has already conducted regarding the subject matter of this Request and its ongoing investigation of other topics. As part of Microsoft's ongoing investigation in this case, it has and will continue to search for (including custodial ESI searches) and produce non-privileged, responsive material related to the training of OpenAI's LLMs.

REQUEST FOR PRODUCTION NO. 42:

Documents Concerning or Relating to any complaints **You** have received regarding **Your** or **OpenAI's** use of potentially copyrighted material in training **Large Language Models**.

RESPONSE TO REQUEST FOR PRODUCTION NO. 42:

Microsoft incorporates by reference and reasserts its General Objections to Plaintiffs' "Definitions" set forth above. Microsoft objects to this Request on the grounds that it is irrelevant and/or overly broad, unduly burdensome, and not proportional to the needs of the case to the extent that it seeks information concerning any complaint of use of any potentially copyrighted material, owned by anyone, at any time, and without any subject matter or geographic limitation. Microsoft further objects to this Request to the extent it assumes that Microsoft trained or otherwise developed OpenAI's models or applications identified in the Consolidated Complaint. Microsoft did not train the Relevant Generative AI Models, does not possess the training datasets OpenAI used to train the Relevant Generative AI Models, and is unaware of the contents of the training datasets used by OpenAI to train the Relevant Generative AI Models. Microsoft further objects to

this Request as irrelevant and/or vague, ambiguous, overly broad, and not proportional to the needs of the case because it requests information concerning all of Microsoft's and OpenAI's training datasets, pre-training datasets, and/or fine-tuning datasets, regardless of whether they have been identified in the Consolidated Complaint and/or bear any relation to the LLMs at issue in these Actions.

Subject to these objections, Microsoft responds as follows: Microsoft will conduct a reasonable search and produce responsive, nonprivileged documents, if any, concerning any complaints of copyright infringement that Microsoft has received regarding OpenAI's use of text data to train Relevant Generative AI Models.

REQUEST FOR PRODUCTION NO. 43:

Documents Concerning or Relating to Your policies or procedures **Relating to** receiving, processing, or responding to complaints regarding **Your** or **OpenAI's** use of potentially copyrighted material in training **Large Language Models**.

RESPONSE TO REQUEST FOR PRODUCTION NO. 43:

Microsoft incorporates by reference and reasserts its General Objections to Plaintiffs' "Definitions" set forth above. Microsoft objects to this Request on the grounds that it is irrelevant and/or overly broad, unduly burdensome, and not proportional to the needs of the case to the extent that it seeks information concerning any complaint of use of any potentially copyrighted material, owned by anyone, at any time, and without any subject matter or geographic limitation. Microsoft further objects to this Request to the extent it assumes that Microsoft trained or otherwise developed OpenAI's models or applications identified in the Consolidated Complaint. Microsoft did not train the Relevant Generative AI Models, does not possess the training datasets OpenAI used to train the Relevant Generative AI Models, and is unaware of the contents of the training datasets used by OpenAI to train the Relevant Generative AI Models. Microsoft further objects to

this Request as irrelevant and/or vague, ambiguous, overly broad, and not proportional to the needs of the case because it requests information concerning all of Microsoft’s and OpenAI’s training datasets, pre-training datasets, and/or fine-tuning datasets, regardless of whether they have been identified in the Consolidated Complaint and/or bear any relation to the LLMs at issue in these Actions.

Subject to these objections, Microsoft responds as follows: Microsoft will conduct a reasonable search and produce responsive, nonprivileged documents, if any, concerning Microsoft’s policies and procedures relating to receiving, processing, or responding to any complaints of copyright infringement that Microsoft has received regarding OpenAI’s use of text data to train Relevant Generative AI Models.

REQUEST FOR PRODUCTION NO. 44:

Documents sufficient to identify any effort by **You** or **OpenAI** to remedy or mitigate **Your** or **OpenAI’s** unauthorized use of copyrighted material to train any **Large Language Models**.

RESPONSE TO REQUEST FOR PRODUCTION NO. 44:

Microsoft incorporates by reference and reasserts its General Objections to Plaintiffs’ “Definitions” set forth above. Microsoft objects to the phrase “remedy or mitigate Your or OpenAI’s unauthorized use of copyrighted material” because it assumes that Microsoft or OpenAI has made unauthorized uses of copyrighted material, which Microsoft does not accept as correct, and which appears to require a legal conclusion. Microsoft also objects to this Request on the grounds that it is irrelevant and/or overly broad, unduly burdensome, and not proportional to the needs of the case to the extent that it seeks information concerning any use of copyrighted material, owned by anyone, at any time, and without any subject matter or geographic limitation. Microsoft further objects to this Request to the extent it assumes that Microsoft trained or otherwise developed OpenAI’s models or applications identified in the Consolidated Complaint. Microsoft

did not train the Relevant Generative AI Models, does not possess the training datasets OpenAI used to train the Relevant Generative AI Models, and is unaware of the contents of the training datasets used by OpenAI to train the Relevant Generative AI Models. Microsoft further objects to this Request as irrelevant and/or vague, ambiguous, overly broad, and not proportional to the needs of the case because it requests information concerning all of Microsoft's and OpenAI's training datasets, pre-training datasets, and/or fine-tuning datasets, regardless of whether they have been identified in the Consolidated Complaint and/or bear any relation to the LLMs at issue in these Actions.

Subject to these objections, Microsoft responds as follows: Because this Request assumes that Microsoft or OpenAI has made unauthorized uses of copyrighted material, which Microsoft does not accept occurred and which appears to require a legal conclusion, Microsoft will not search for and produce documents in response to this Request.

REQUEST FOR PRODUCTION NO. 45:

All **Communications** between January 1, 2016 and December 31, 2022, related to the use or reproduction of copyrighted materials to train any artificial intelligence model **OpenAI** developed, including without limitation all **Large Language Models** and diffusion models (e.g., DALL-E, Codex, GPT-2).

RESPONSE TO REQUEST FOR PRODUCTION NO. 45:

Microsoft incorporates by reference and reasserts its General Objections to Plaintiffs' "Definitions" set forth above. Microsoft objects to the timeframe of January 1, 2016, to December 31, 2022, on the grounds that it is irrelevant and/or overly broad, unduly burdensome, and not proportional to the needs of the case to the extent that it seeks communications prior to when Microsoft partnered with OpenAI in 2019. Microsoft also objects to the term "reproduction" to the extent it assumes or implies that Microsoft or OpenAI "reproduce[ed]" any "copyrighted

materials” as defined by the Copyright Act. Microsoft also objects to the Request as overly broad, unduly burdensome, and not proportional to the needs of the case to the extent it seeks “[a]ll Communications.” Microsoft further objects to this Request on the grounds that it is irrelevant and/or overly broad, unduly burdensome, and not proportional to the needs of the case to the extent that it seeks information concerning any use of any copyrighted material, owned by anyone, at any time, and without any subject matter or geographic limitation. Microsoft further objects to this Request to the extent it assumes that Microsoft trained or otherwise developed OpenAI’s models or applications identified in the Consolidated Complaint. Microsoft did not train the Relevant Generative AI Models, does not possess the training datasets OpenAI used to train the Relevant Generative AI Models, and is unaware of the contents of the training datasets used by OpenAI to train the Relevant Generative AI Models. Microsoft further objects to this Request as irrelevant and/or vague, ambiguous, overly broad, and not proportional to the needs of the case because it requests information concerning all of Microsoft’s and OpenAI’s training datasets, pre-training datasets, and/or fine-tuning datasets, regardless of whether they have been identified in the Consolidated Complaint and/or bear any relation to the LLMs at issue in these Actions.

Subject to these objections, Microsoft responds as follows: Microsoft objects to this Request because it implies, suggests, and/or assumes that Microsoft trained or otherwise developed OpenAI’s models. Microsoft has already made a thorough investigation (including custodial and non-custodial material) into the issue of whether it had any role in OpenAI’s pretraining of its LLMs. As such, Microsoft is relying upon this objection for the purpose of limiting the scope of its investigation on this particular topic to the reasonable search it has already conducted regarding the subject matter of this Request and its ongoing investigation of other topics. As part of Microsoft’s ongoing investigation in this case, it has and will continue to search for (including

custodial ESI searches) and produce non-privileged, responsive material related to the training of OpenAI's LLMs.

REQUEST FOR PRODUCTION NO. 46:

Documents sufficient to show all attempts by **You** or **OpenAI** to obtain a license for the use of text or audio works in connection with the development of any **Large Language Model(s)**, including but not limited to **Documents** reflecting any negotiations of such licenses.

RESPONSE TO REQUEST FOR PRODUCTION NO. 46:

Microsoft incorporates by reference and reasserts its General Objections to Plaintiffs' "Definitions" set forth above. Microsoft objects to this Request on the grounds that it is irrelevant and/or overly broad, unduly burdensome, and not proportional to the needs of the case to the extent that it seeks information concerning licensing of any works, owned by anyone, at any time, and without any subject matter or geographic limitation. Microsoft also objects to this Request to the extent that it includes "audio works" on the grounds that is irrelevant and/or overly broad, unduly burdensome, and not proportional to the needs of the case in that the Consolidated Complaint does not include any allegations regarding audio works. Microsoft also objects to the term "development" as vague and ambiguous in that this term is undefined. Microsoft understands "development" as meaning "training" in this Request. Microsoft further objects to the term "license" as vague and ambiguous. Microsoft will interpret the term "license" to mean "an agreement to access or use." Microsoft also objects to this Request as irrelevant and/or overly broad and not proportional to the needs of the case because it requests information concerning "attempts" to obtain a license. Microsoft will produce documents concerning licensing agreements that were completed, finalized, and/or executed. Microsoft further objects to this Request to the extent it assumes that Microsoft trained or otherwise developed OpenAI's models or applications identified in the Consolidated Complaint. Microsoft did not train the Relevant Generative AI

Models, does not possess the training datasets OpenAI used to train the Relevant Generative AI Models, and is unaware of the contents of the training datasets used by OpenAI to train the Relevant Generative AI Models.

Subject to these objections, Microsoft responds as follows: Microsoft will conduct a reasonable search and, pending fulfillment of any notification or confidentiality obligations Microsoft may have to applicable third-party individuals or organizations, produce responsive, nonprivileged documents, if any, sufficient to show licensing and/or data-access agreements in the United States, dated on or after January 1, 2019, that provide for Microsoft and/or OpenAI to receive, access, and/or use text data in training Large Language Models. For the purposes of this response only, “Large Language Models” refers to deep learning models that are trained using very large amounts of text data to generate text, including but not limited to LLMs as defined in Microsoft’s General Objections.

REQUEST FOR PRODUCTION NO. 47:

All **Documents Concerning** or **Relating to Your** decision(s) to seek or obtain a license to use any copyrighted material as **Training Material**.

RESPONSE TO REQUEST FOR PRODUCTION NO. 47:

Microsoft incorporates by reference and reasserts its General Objections to Plaintiffs’ “Definitions” set forth above. Microsoft objects to this Request on the grounds that it is irrelevant and/or overly broad, unduly burdensome, and not proportional to the needs of the case to the extent that it seeks information concerning licensing of any copyrighted material, owned by anyone, at any time, and without any subject matter or geographic limitation. Microsoft also objects to this Request to the extent that it includes “any” material on the grounds that is irrelevant and/or overly broad, unduly burdensome, and not proportional to the needs of the case in that the Consolidated Complaint alleges conduct related to text works and not “any” material. Microsoft objects to the

phrase “decision(s) to seek or obtain a license” as vague and ambiguous because it assumes that Microsoft made or was otherwise involved with any decision, in any capacity, including the consideration, contemplation, or effectuation thereof, regarding the contents of the data used by OpenAI for training. Microsoft also objects to the Request as overly broad, unduly burdensome, and not proportional to the needs of the case to the extent it seeks “[a]ll Documents.” Microsoft further objects to the term “license” as vague and ambiguous. Microsoft will interpret the term “license” to mean “an agreement to access or use.” Microsoft also objects to this Request as irrelevant and/or overly broad and not proportional to the needs of the case because it requests information concerning “decision(s) to seek or obtain” a license. Microsoft will produce documents concerning licensing agreements that were completed, finalized, and/or executed. Microsoft further objects to this Request to the extent it assumes that Microsoft trained or otherwise developed OpenAI’s models or applications identified in the Consolidated Complaint. Microsoft did not train the Relevant Generative AI Models, does not possess the training datasets OpenAI used to train the Relevant Generative AI Models, and is unaware of the contents of the training datasets used by OpenAI to train the Relevant Generative AI Models.

Subject to these objections, Microsoft responds as follows: Microsoft will conduct a reasonable search and, pending fulfillment of any notification or confidentiality obligations Microsoft may have to applicable third-party individuals or organizations, produce responsive, nonprivileged documents, if any, concerning Microsoft’s licensing and/or data-access agreements in the United States, dated on or after January 1, 2019, that provide for Microsoft and/or OpenAI to receive, access, and/or use text data in training Large Language Models. For the purposes of this response only, “Large Language Models” refers to deep learning models that are trained using

very large amounts of text data to generate text, including but not limited to LLMs as defined in Microsoft's General Objections.

REQUEST FOR PRODUCTION NO. 48:

All Documents Concerning or Relating to the relative value of different types of **Training Material.**

RESPONSE TO REQUEST FOR PRODUCTION NO. 48:

Microsoft incorporates by reference and reasserts its General Objections to Plaintiffs' "Definitions" set forth above. Microsoft objects to the phrase "different types" as vague and ambiguous in that this phrase is undefined. Microsoft also objects to the Request as overly broad, unduly burdensome, and not proportional to the needs of the case to the extent it seeks "[a]ll Documents." Microsoft further objects to this Request to the extent it assumes that Microsoft trained or otherwise developed OpenAI's models or applications identified in the Consolidated Complaint. Microsoft did not train the Relevant Generative AI Models, does not possess the training datasets OpenAI used to train the Relevant Generative AI Models, and is unaware of the contents of the training datasets used by OpenAI to train the Relevant Generative AI Models. Microsoft further objects to this Request as irrelevant and/or vague, ambiguous, overly broad, and not proportional to the needs of the case because it requests information concerning all of Microsoft's and OpenAI's training datasets, pre-training datasets, and/or fine-tuning datasets, regardless of whether they have been identified in the Consolidated Complaint and/or bear any relation to the LLMs at issue in these Actions.

Subject to these objections, Microsoft responds as follows: Microsoft will conduct a reasonable search and produce responsive, nonprivileged documents, if any, concerning the value of text data in connection with OpenAI's training of the Relevant Generative AI Models.

REQUEST FOR PRODUCTION NO. 49:

All **Documents** concerning or relating to **OpenAI** deriving **Training Material** from any physical books or audiobooks.

RESPONSE TO REQUEST FOR PRODUCTION NO. 49:

Microsoft incorporates by reference and reasserts its General Objections to Plaintiffs’ “Definitions” set forth above. Microsoft objects to this Request on the grounds that it is irrelevant and/or overly broad, unduly burdensome, and not proportional to the needs of the case to the extent that it seeks information concerning use of any physical books or audiobooks, owned by anyone, at any time, and without any subject matter or geographic limitation. Microsoft also objects to this Request as vague and ambiguous because the meaning of “deriving” in this Request is unclear and objects to this term to the extent it assumes or implies that any Training Material is a derivative work as defined by the Copyright Act. Microsoft also objects to the Request as overly broad, unduly burdensome, and not proportional to the needs of the case to the extent it seeks “[a]ll Documents.” Microsoft further objects to this Request to the extent it assumes that Microsoft trained or otherwise developed OpenAI’s models or applications identified in the Consolidated Complaint. Microsoft did not train the Relevant Generative AI Models, does not possess the training datasets OpenAI used to train the Relevant Generative AI Models, and is unaware of the contents of the training datasets used by OpenAI to train the Relevant Generative AI Models. Microsoft also objects to this Request to the extent that it includes “audiobooks” on the grounds that is irrelevant and/or overly broad, unduly burdensome, and not proportional to the needs of the case in that the Consolidated Complaint does not include any allegations regarding audiobooks. Microsoft further objects to this Request as irrelevant and/or vague, ambiguous, overly broad, and not proportional to the needs of the case because it requests information concerning all of OpenAI’s training datasets, pre-training datasets, and/or fine-tuning datasets, regardless of whether they have

been identified in the Consolidated Complaint and/or bear any relation to the LLMs at issue in these Actions.

Subject to these objections, Microsoft responds as follows: Microsoft objects to this Request because it implies, suggests, and/or assumes that Microsoft trained or otherwise developed OpenAI's models. Microsoft has already made a thorough investigation (including custodial and non-custodial material) into the issue of whether it had any role in OpenAI's pretraining of its LLMs. As such, Microsoft is relying upon this objection for the purpose of limiting the scope of its investigation on this particular topic to the reasonable search it has already conducted regarding the subject matter of this Request and its ongoing investigation of other topics. As part of Microsoft's ongoing investigation in this case, it has and will continue to search for (including custodial ESI searches) and produce non-privileged, responsive material related to the training of OpenAI's LLMs.

REQUEST FOR PRODUCTION NO. 50:

All Documents Concerning or Relating to OpenAI's use of digital books as Training Material.

RESPONSE TO REQUEST FOR PRODUCTION NO. 50:

Microsoft incorporates by reference and reasserts its General Objections to Plaintiffs' "Definitions" set forth above. Microsoft objects to this Request on the grounds that it is irrelevant and/or overly broad, unduly burdensome, and not proportional to the needs of the case to the extent that it seeks information concerning use of any digital books, owned by anyone, at any time, and without any subject matter or geographic limitation. Microsoft also objects to the Request as overly broad, unduly burdensome, and not proportional to the needs of the case to the extent it seeks "[a]ll Documents." Microsoft further objects to this Request to the extent it assumes that Microsoft trained or otherwise developed OpenAI's models or applications identified in the Consolidated

Complaint. Microsoft did not train the Relevant Generative AI Models, does not possess the training datasets OpenAI used to train the Relevant Generative AI Models, and is unaware of the contents of the training datasets used by OpenAI to train the Relevant Generative AI Models. Microsoft further objects to this Request as irrelevant and/or vague, ambiguous, overly broad, and not proportional to the needs of the case because it requests information concerning all of OpenAI's training datasets, pre-training datasets, and/or fine-tuning datasets, regardless of whether they have been identified in the Consolidated Complaint and/or bear any relation to the LLMs at issue in these Actions.

Subject to these objections, Microsoft responds as follows: Microsoft objects to this Request because it implies, suggests, and/or assumes that Microsoft trained or otherwise developed OpenAI's models. Microsoft has already made a thorough investigation (including custodial and non-custodial material) into the issue of whether it had any role in OpenAI's pretraining of its LLMs. As such, Microsoft is relying upon this objection for the purpose of limiting the scope of its investigation on this particular topic to the reasonable search it has already conducted regarding the subject matter of this Request and its ongoing investigation of other topics. As part of Microsoft's ongoing investigation in this case, it has and will continue to search for (including custodial ESI searches) and produce non-privileged, responsive material related to the training of OpenAI's LLMs.

REQUEST FOR PRODUCTION NO. 51:

Documents sufficient to show any reproduction of the data **OpenAI** used to train any **Large Language Model**, including any reproductions that came into Microsoft's possession.

RESPONSE TO REQUEST FOR PRODUCTION NO. 51:

Microsoft incorporates by reference and reasserts its General Objections to Plaintiffs' "Definitions" set forth above. Microsoft objects to the term "reproduction" to the extent it assumes

or implies that Microsoft or OpenAI “reproduc[ed]” any data as defined by the Copyright Act. Microsoft also objects to this Request on the grounds that it is irrelevant and/or overly broad, unduly burdensome, and not proportional to the needs of the case to the extent that it seeks information concerning use of any data, owned by anyone, at any time, and without any subject matter or geographic limitation. Microsoft further objects to this Request to the extent it assumes that Microsoft trained or otherwise developed OpenAI’s models or applications identified in the Consolidated Complaint. Microsoft did not train the Relevant Generative AI Models, does not possess the training datasets OpenAI used to train the Relevant Generative AI Models, and is unaware of the contents of the training datasets used by OpenAI to train the Relevant Generative AI Models. Microsoft further objects to this Request as irrelevant and/or vague, ambiguous, overly broad, and not proportional to the needs of the case because it requests information concerning all of OpenAI’s training datasets, pre-training datasets, and/or fine-tuning datasets, regardless of whether they have been identified in the Consolidated Complaint and/or bear any relation to the LLMs at issue in these Actions.

Subject to these objections, Microsoft will conduct a reasonable search and produce responsive, nonprivileged documents, if any, concerning any reproduction of the data OpenAI used to train any Relevant Generative AI Models that came into Microsoft’s possession.

REQUEST FOR PRODUCTION NO. 52:

All **Documents** and **Communications Concerning** or **Relating** to the use or contemplated use of material from any text data repository (e.g. Library Genesis, Internet Archive, Z-Library, Common Crawl, WebText, Project Gutenberg, Anna’s Archive, Open Library, Reddit, DuXiu) as **Training Material**.

RESPONSE TO REQUEST FOR PRODUCTION NO. 52:

Microsoft incorporates by reference and reasserts its General Objections to Plaintiffs’ “Definitions” set forth above. Microsoft objects to this Request on the grounds that it is irrelevant and/or overly broad, unduly burdensome, and not proportional to the needs of the case to the extent that it seeks information concerning use or contemplated use of material from any text data repository, owned by anyone, at any time, and without any subject matter or geographic limitation. Microsoft also objects to the Request as overly broad, unduly burdensome, and not proportional to the needs of the case to the extent it seeks “[a]ll Documents and Communications.” Microsoft further objects to this Request to the extent it assumes that Microsoft trained or otherwise developed OpenAI’s models or applications identified in the Consolidated Complaint. Microsoft did not train the Relevant Generative AI Models, does not possess the training datasets OpenAI used to train the Relevant Generative AI Models, and is unaware of the contents of the training datasets used by OpenAI to train the Relevant Generative AI Models. Microsoft further objects to this Request as irrelevant and/or vague, ambiguous, overly broad, and not proportional to the needs of the case because it requests information concerning all of Microsoft’s and OpenAI’s training datasets, pre-training datasets, and/or fine-tuning datasets, regardless of whether they have been identified in the Consolidated Complaint and/or bear any relation to the LLMs at issue in these Actions.

Subject to these objections, Microsoft responds as follows: Microsoft objects to this Request because it implies, suggests, and/or assumes that Microsoft trained or otherwise developed OpenAI’s models. Microsoft has already made a thorough investigation (including custodial and non-custodial material) into the issue of whether it had any role in OpenAI’s pretraining of its LLMs. As such, Microsoft is relying upon this objection for the purpose of limiting the scope of

its investigation on this particular topic to the reasonable search it has already conducted regarding the subject matter of this Request and its ongoing investigation of other topics. As part of Microsoft's ongoing investigation in this case, it has and will continue to search for (including custodial ESI searches) and produce non-privileged, responsive material related to the training of OpenAI's LLMs.

REQUEST FOR PRODUCTION NO. 53:

All **Documents Concerning or Relating to OpenAI's** potential exhaustion of available text data for use as **Training Material**.

RESPONSE TO REQUEST FOR PRODUCTION NO. 53:

Microsoft incorporates by reference and reasserts its General Objections to Plaintiffs' "Definitions" set forth above. Microsoft objects to this Request on the grounds that it is irrelevant and/or overly broad, unduly burdensome, and not proportional to the needs of the case to the extent that it seeks information concerning available text data, owned by anyone, at any time, and without any subject matter or geographic limitation. Microsoft also objects to this Request as vague, and ambiguous because "exhaustion of available text data" is not defined. Microsoft understands this phrase to mean "lack of new text data and/or different text data from data already used as Training Material." Microsoft also objects to the Request as overly broad, unduly burdensome, and not proportional to the needs of the case to the extent it seeks "[a]ll Documents." Microsoft further objects to this Request to the extent it assumes that Microsoft trained or otherwise developed OpenAI's models or applications identified in the Consolidated Complaint. Microsoft did not train the Relevant Generative AI Models, does not possess the training datasets OpenAI used to train the Relevant Generative AI Models, and is unaware of the contents of the training datasets used by OpenAI to train the Relevant Generative AI Models. Microsoft further objects to this Request as irrelevant and/or vague, ambiguous, overly broad, and not proportional to the needs of the case

because it requests information concerning all of OpenAI’s training datasets, pre-training datasets, and/or fine-tuning datasets, regardless of whether they have been identified in the Consolidated Complaint and/or bear any relation to the LLMs at issue in these Actions.

Subject to these objections, Microsoft responds as follows: Microsoft will conduct a reasonable search and produce responsive, nonprivileged documents, if any, concerning OpenAI’s potential lack of new text data and/or different text data from data already used as Training Material.

REQUEST FOR PRODUCTION NO. 54:

Documents sufficient to show when **You** became aware of each dataset used to train **OpenAI’s Large Language Models**, including but not limited to, all **Documents Concerning** and **Relating to** the “writeup of OpenAI’s technical results on large language models” **OpenAI** sent to **You** in February 2020 that “include[d] details on training methods, datasets, etc to facilitate comparisons and collaboration.” (MSFT_AICPY_000228653).

RESPONSE TO REQUEST FOR PRODUCTION NO. 54:

Microsoft incorporates by reference and reasserts its General Objections to Plaintiffs’ “Definitions” set forth above. Microsoft objects to this Request because the phrase “became aware of each dataset used to train OpenAI’s Large Language Models” assumes that Microsoft gained such awareness, which Microsoft does not accept occurred. Microsoft further objects to this Request to the extent it assumes that Microsoft trained or otherwise developed OpenAI’s models or applications identified in the Consolidated Complaint. Microsoft did not train the Relevant Generative AI Models, does not possess the training datasets OpenAI used to train the Relevant Generative AI Models, and is unaware of the contents of the training datasets used by OpenAI to train the Relevant Generative AI Models. Microsoft objects to this Request as irrelevant and/or vague, ambiguous, overly broad, and not proportional to the needs of the case because it requests

information concerning all of OpenAI’s training datasets, pre-training datasets, and/or fine-tuning datasets, regardless of whether they have been identified in the Consolidated Complaint and/or bear any relation to the LLMs at issue in these Actions.

Subject to these objections, Microsoft responds as follows: Microsoft objects to this Request because it implies, suggests, and/or assumes that Microsoft trained or otherwise developed OpenAI’s models. Microsoft has already made a thorough investigation (including custodial and non-custodial material) into the issue of whether it had any role in OpenAI’s pretraining of its LLMs. As such, Microsoft is relying upon this objection for the purpose of limiting the scope of its investigation on this particular topic to the reasonable search it has already conducted regarding the subject matter of this Request and its ongoing investigation of other topics. As part of Microsoft’s ongoing investigation in this case, it has and will continue to search for (including custodial ESI searches) and produce non-privileged, responsive material related to the training of OpenAI’s LLMs.

REQUEST FOR PRODUCTION NO. 55:

All Documents Concerning or Relating to the relative value of different types of text data in training Large Language Models.

RESPONSE TO REQUEST FOR PRODUCTION NO. 55:

Microsoft incorporates by reference and reasserts its General Objections to Plaintiffs’ “Definitions” set forth above. Microsoft objects to this Request as duplicative of Request No. 48. Microsoft also objects to the phrase “different types” as vague and ambiguous in that this phrase is undefined. Microsoft also objects to the Request as overly broad, unduly burdensome, and not proportional to the needs of the case to the extent it seeks “[a]ll Documents.” Microsoft further objects to this Request to the extent it assumes that Microsoft trained or otherwise developed OpenAI’s models or applications identified in the Consolidated Complaint. Microsoft did not train

the Relevant Generative AI Models, does not possess the training datasets OpenAI used to train the Relevant Generative AI Models, and is unaware of the contents of the training datasets used by OpenAI to train the Relevant Generative AI Models. Microsoft further objects to this Request as irrelevant and/or vague, ambiguous, overly broad, and not proportional to the needs of the case because it requests information concerning all of Microsoft's and OpenAI's training datasets, pre-training datasets, and/or fine-tuning datasets, regardless of whether they have been identified in the Consolidated Complaint and/or bear any relation to the LLMs at issue in these Actions.

Subject to these objections, Microsoft responds as follows: Microsoft will conduct a reasonable search and produce responsive, nonprivileged documents, if any, concerning the value of text data in connection with OpenAI's training of the Relevant Generative AI Models.

REQUEST FOR PRODUCTION NO. 56:

All **Documents Concerning** or **Relating to** any programming of any **Large Language Models** developed by **OpenAI** designed or intended—in whole or in part—to avoid copyright infringement and/or to avoid reproducing materials used to train the **Large Language Models** or **ChatGPT**.

RESPONSE TO REQUEST FOR PRODUCTION NO. 56:

Microsoft incorporates by reference and reasserts its General Objections to Plaintiffs' "Definitions" set forth above. Microsoft objects to this Request on the grounds that it is irrelevant and/or overly broad, unduly burdensome, and not proportional to the needs of the case to the extent it seeks information that is not relevant to the Consolidated Complaint, which focus on use of Plaintiffs' works to train the LLMs, not development or use of an LLM after training or outputs of products or applications using an LLM. Microsoft also objects to the Request as vague and ambiguous because the meaning of "programming" in this Request is unclear. Microsoft also objects to the Request as overly broad, unduly burdensome, and not proportional to the needs of

the case to the extent it seeks “[a]ll Documents and Communications.” Microsoft further objects to this Request to the extent it assumes that Microsoft programmed, trained, or otherwise developed OpenAI’s models or applications identified in the Consolidated Complaint. Microsoft did not train the Relevant Generative AI Models, does not possess the training datasets OpenAI used to train the Relevant Generative AI Models, and is unaware of the contents of the training datasets used by OpenAI to train the Relevant Generative AI Models, and further does not develop ChatGPT and/or OpenAI’s consumer-facing applications. Microsoft further objects to this Request as irrelevant and/or vague, ambiguous, overly broad, and not proportional to the needs of the case because it requests information concerning all of OpenAI’s LLMs, regardless of whether they have been identified in the Consolidated Complaint and/or bear any relation to the LLMs at issue in these Actions.

Subject to these objections, Microsoft responds as follows: Microsoft will not currently search for and produce documents in response to this Request, but is willing to meet and confer regarding the scope of this Request.

REQUEST FOR PRODUCTION NO. 57:

All **Documents Concerning or Relating to** the impact that including or excluding human-generated text from a **Large Language Model’s** training data has on the quality of its outputs.

RESPONSE TO REQUEST FOR PRODUCTION NO. 57:

Microsoft incorporates by reference and reasserts its General Objections to Plaintiffs’ “Definitions” set forth above. Microsoft also objects to the Request as overly broad, unduly burdensome, and not proportional to the needs of the case to the extent it seeks “[a]ll Documents.” Microsoft further objects to this Request to the extent it assumes that Microsoft programmed, trained, or otherwise developed OpenAI’s models or applications identified in the Consolidated Complaint. Microsoft did not train the Relevant Generative AI Models, does not possess the

training datasets OpenAI used to train the Relevant Generative AI Models, and is unaware of the contents of the training datasets used by OpenAI to train the Relevant Generative AI Models. Microsoft further objects to this Request as irrelevant and/or vague, ambiguous, overly broad, and not proportional to the needs of the case because it requests information concerning all of Microsoft's and OpenAI's training datasets, pre-training datasets, and/or fine-tuning datasets, regardless of whether they have been identified in the Consolidated Complaint and/or bear any relation to the LLMs at issue in these Actions.

Subject to these objections, Microsoft responds as follows: Microsoft will conduct a reasonable search and produce responsive, nonprivileged documents, if any, concerning the impact that including or excluding human-generated text from the Relevant Generative AI Models has on the quality of its outputs.

REQUEST FOR PRODUCTION NO. 58:

Documents sufficient to show any exclusive or non-exclusive rights granted to **You** by **OpenAI** to commercialize, monetize, or productize **OpenAI's Large Language Models**, training data, or intellectual property.

RESPONSE TO REQUEST FOR PRODUCTION NO. 58:

Microsoft incorporates by reference and reasserts its General Objections to Plaintiffs' "Definitions" Microsoft objects to the Request as vague and ambiguous because the term "productize" is undefined. Microsoft objects to this Request as irrelevant and/or vague, ambiguous, overly broad, and not proportional to the needs of the case because it requests information concerning all of OpenAI's training datasets, pre-training datasets, and/or fine-tuning datasets, regardless of whether they have been identified in the Consolidated Complaint and/or bear any relation to the LLMs at issue in these Actions.

Subject to these objections, Microsoft responds as follows: Microsoft has produced the agreements between it and OpenAI related to Microsoft's rights as to the Relevant Generative AI Models and the intellectual property related to the Relevant Generative AI Models.

REQUEST FOR PRODUCTION NO. 59:

Documents sufficient to identify all of Your commercial products that use or include **any Large Language Model developed by OpenAI**, including but not limited to, **Your** Bing search engine (as set forth in the First Consolidated Complaint [Fiction Docket No. 69] at ¶166) as well as Microsoft Teams, Microsoft 365 Copilot and GitHub Copilot (*id.* at ¶167).

RESPONSE TO REQUEST FOR PRODUCTION NO. 59:

Microsoft incorporates by reference and reasserts its General Objections to Plaintiffs' "Definitions" Microsoft objects to this Request as irrelevant and/or vague, ambiguous, overly broad, and not proportional to the needs of the case because it requests information concerning all of OpenAI's LLMs, regardless of whether they have been identified in the Consolidated Complaint and/or bear any relation to the LLMs at issue in these Actions. Microsoft further objects to this Request as irrelevant and/or vague, ambiguous, overly broad, and not proportional to the needs of the case because it seeks information related to commercial products that bear no relationship to Plaintiffs' infringement allegations in the Consolidated Complaint, which focus on training of the Relevant Generative AI Models.

Microsoft responds as follows: Microsoft will not currently search for documents responsive to this Request but is willing to meet and confer regarding the scope of this Request.

REQUEST FOR PRODUCTION NO. 60:

Documents sufficient to identify each person within **Microsoft** who participated, was involved in or was consulted **Concerning** the use, implementation or incorporation of **ChatGPT** or **Large Language Models** in **Your** commercial products.

RESPONSE TO REQUEST FOR PRODUCTION NO. 60:

Microsoft incorporates by reference and reasserts its General Objections to Plaintiffs’ “Definitions” Microsoft objects to this Request as irrelevant and/or vague, ambiguous, overly broad, and not proportional to the needs of the case because it requests information concerning all of LLMs and versions of ChatGPT utilized in Microsoft products, regardless of whether they have been identified in the Consolidated Complaint and/or bear any relation to the LLMs at issue in these Actions. Microsoft objects to this Request as irrelevant and/or vague, ambiguous, overly broad, and not proportional to the needs of the case because it requests information concerning “each” person where identifying “each” such person is not relevant nor proportionate to the case. Microsoft also objects to this Request as irrelevant and/or overly broad and not proportional to the needs of the case because it seeks information related to Microsoft’s development of products incorporating LLMs, which is not relevant to allegations in the Consolidated Complaint. Microsoft further objects to this Request as irrelevant and/or vague, ambiguous, overly broad, and not proportional to the needs of the case because it seeks information related to commercial products that bear no relationship to Plaintiffs’ infringement allegations in the Consolidated Complaint, which focus on training of the Relevant Generative AI Models.

Microsoft responds as follows: Microsoft will not currently search for documents responsive to this Request but is willing to meet and confer regarding the scope of this Request.

REQUEST FOR PRODUCTION NO. 61:

All **Documents Concerning or Relating to** customers substituting away from or towards a **Microsoft** product using a **Large Language Model** based on the quality of the **Large Language Model**.

RESPONSE TO REQUEST FOR PRODUCTION NO.61:

Microsoft incorporates by reference and reasserts its General Objections to Plaintiffs’ “Definitions” set forth above. Microsoft also objects to the Request as overly broad, unduly burdensome, and not proportional to the needs of the case to the extent it seeks “[a]ll Documents.” Microsoft also objects to the phrase “substituting away from or towards” as vague and ambiguous in that this phrase is undefined. Microsoft further objects to this Request as irrelevant and/or vague, ambiguous, overly broad, and not proportional to the needs of the case because it requests information concerning all LLMs utilized in Microsoft products, regardless of whether they have been identified in the Consolidated Complaint and/or bear any relation to the LLMs at issue in these Actions. Microsoft further objects to this Request as irrelevant and/or vague, ambiguous, overly broad, and not proportional to the needs of the case because it seeks information related to Microsoft products that bear no relationship to Plaintiffs’ infringement allegations in the Consolidated Complaint, which focus on training of the Relevant Generative AI Models.

Microsoft responds as follows: Microsoft will not currently search for documents responsive to this Request but is willing to meet and confer regarding the scope of this Request.

REQUEST FOR PRODUCTION NO. 62:

Documents sufficient to identify **Your** gross revenues, net revenues, and profits, by month, from Your commercial products that include or use **ChatGPT** or **Large Language Models**.

RESPONSE TO REQUEST FOR PRODUCTION NO. 62:

Microsoft incorporates by reference and reasserts its General Objections to Plaintiffs’ “Definitions” set forth above. Microsoft further objects to this Request as irrelevant and/or vague, ambiguous, overly broad, and not proportional to the needs of the case because it requests information concerning all LLMs utilized in Microsoft products, regardless of whether they have been identified in the Consolidated Complaint and/or bear any relation to the LLMs at issue in

these Actions. Microsoft also objects to this Request to the extent it seeks information on revenues, which are not relevant to the allegations in the Consolidated Complaint. Microsoft further objects to this Request as irrelevant and/or vague, ambiguous, overly broad, and not proportional to the needs of the case because it seeks information related to commercial products that bear no relationship to Plaintiffs' infringement allegations in the Consolidated Complaint, which focus on training of the Relevant Generative AI Models.

Microsoft responds as follows: Microsoft will not currently search for documents responsive to this Request but is willing to meet and confer regarding the scope of this Request.

REQUEST FOR PRODUCTION NO. 63:

Documents sufficient to show the gross revenues, net revenues, and profits, by month, generated by each **Large Language Model You** have commercialized, sold, and/or licensed, either as a standalone product or as part of another product.

RESPONSE TO REQUEST FOR PRODUCTION NO. 63:

Microsoft incorporates by reference and reasserts its General Objections to Plaintiffs' "Definitions" set forth above. Microsoft further objects to this Request as irrelevant and/or vague, ambiguous, overly broad, and not proportional to the needs of the case because it requests information concerning all LLMs utilized in Microsoft products, regardless of whether they have been identified in the Consolidated Complaint and/or bear any relation to the LLMs at issue in these Actions. Microsoft further objects to this Request as irrelevant and/or vague, ambiguous, overly broad, and not proportional to the needs of the case because it seeks information related to commercial products that bear no relationship to Plaintiffs' infringement allegations in the Consolidated Complaint, which focus on training of the Relevant Generative AI Models.

Microsoft responds as follows: Microsoft will not currently search for documents responsive to this Request but is willing to meet and confer regarding the scope of this Request.

REQUEST FOR PRODUCTION NO. 64:

Document sufficient to show Your balance sheet, income statement, and cash flow statement, on a monthly basis during the relevant time period.

RESPONSE TO REQUEST FOR PRODUCTION NO. 64:

Microsoft incorporates by reference and reasserts its General Objections to Plaintiffs’ “Definitions” set forth above. Microsoft further objects to this Request as irrelevant and/or vague, ambiguous, overly broad, and not proportional to the needs of the case because it requests financial information that is not tailored to and/or does not bear any relation to the allegations in the Consolidated Complaint.

Microsoft responds as follows: Microsoft will not search for and produce documents responsive to this Request.

REQUEST FOR PRODUCTION NO. 65:

All **Documents Concerning or Relating to** predictions, forecasts, or projections of profits, revenues or cash flows of **OpenAI** or from **Large Language Models** or products using or containing **Large Language Models**.

RESPONSE TO REQUEST FOR PRODUCTION NO. 65:

Microsoft incorporates by reference and reasserts its General Objections to Plaintiffs’ “Definitions” set forth above. Microsoft also objects to this Request as overly broad, unduly burdensome, and not proportional to the needs of the case to the extent it seeks “[a]ll Documents.” Microsoft also objects to the Request to the extent that it seeks information related to predictions, forecasts, or projections of profits, revenues, or cash flows of OpenAI as requesting documents that are not in Microsoft’s possession, custody, or control. Microsoft also objects to this Request to the extent it seeks information regarding predictions, forecasts, and projections, which are not relevant to the allegations in the Consolidated Complaint. Microsoft further objects to this Request

as irrelevant and/or vague, ambiguous, overly broad, and not proportional to the needs of the case because it requests information concerning all LLMs utilized in Microsoft products, regardless of whether they have been identified in the Consolidated Complaint and/or bear any relation to the LLMs at issue in these Actions. Microsoft further objects to this Request as irrelevant and/or vague, ambiguous, overly broad, and not proportional to the needs of the case because it seeks information related to commercial products or services that bear no relationship to Plaintiffs' infringement allegations in the Consolidated Complaint, which focus on training of the Relevant Generative AI Models.

Microsoft responds as follows: Microsoft will conduct a reasonable search and produce responsive, nonprivileged documents, if any, concerning predictions, forecasts, or projections of profits, revenues, or cash flows of OpenAI and is willing to meet and confer regarding the scope of this Request.

REQUEST FOR PRODUCTION NO. 66:

All presentations, memoranda, or other **Documents** provided to **Your** Board of Directors **Concerning or Relating to OpenAI's** development of Large Language Models.

RESPONSE TO REQUEST FOR PRODUCTION NO. 66:

Microsoft incorporates by reference and reasserts its General Objections to Plaintiffs' "Definitions" set forth above. Microsoft also objects to this Request as overly broad, unduly burdensome, and not proportional to the needs of the case to the extent it seeks "[a]ll Documents." Microsoft further objects to this Request as irrelevant and/or vague, ambiguous, overly broad, and not proportional to the needs of the case because it requests information concerning all LLMs, regardless of whether they have been identified in the Consolidated Complaint and/or bear any relation to the LLMs at issue in these Actions.

Subject to these objections, Microsoft responds as follows: Microsoft has produced documents responsive to this Request and will produce responsive, nonprivileged documents that have not yet been produced, if any, provided to Microsoft's Board of Directors relating to OpenAI's development of Relevant Generative AI Models.

REQUEST FOR PRODUCTION NO. 67:

Documents sufficient to show each version of the terms and/or conditions of use of **Your** products that use or include **ChatGPT** or **Large Language Models**.

RESPONSE TO REQUEST FOR PRODUCTION NO. 67:

Microsoft incorporates by reference and reasserts its General Objections to Plaintiffs' "Definitions" set forth above. Microsoft further objects to this Request as irrelevant and/or vague, ambiguous, overly broad, and not proportional to the needs of the case because it requests information concerning all LLMs utilized in Microsoft products, regardless of whether they have been identified in the Consolidated Complaint and/or bear any relation to the LLMs at issue in these Actions. Microsoft further objects to this Request as irrelevant and/or vague, ambiguous, overly broad, and not proportional to the needs of the case because it seeks information related to commercial products that bear no relationship to Plaintiffs' infringement allegations in the Consolidated Complaint, which focus on training of the Relevant Generative AI Models.

Microsoft responds as follows: Microsoft will not currently search for documents responsive to this Request but is willing to meet and confer regarding the scope of this Request.

REQUEST FOR PRODUCTION NO. 68:

Retention agreements You entered from January 1, 2016 to the present, with any law firm(s) for actual or potential copyright litigation **Related to Your** investment in and/or partnership with **OpenAI**.

RESPONSE TO REQUEST FOR PRODUCTION NO. 68:

Microsoft incorporates by reference and reasserts its General Objections to Plaintiffs’ “Definitions” set forth above. Microsoft further objects to this Request as irrelevant and/or vague, ambiguous, overly broad, and not proportional to the needs of the case because it requests information that does not bear any relation to the allegations in the Consolidated Complaint. Microsoft also objects to the timeframe of January 1, 2016, to present on the grounds that it is irrelevant and/or overly broad, unduly burdensome, and not proportional to the needs of the case to the extent that it seeks communications prior to when Microsoft partnered with OpenAI in 2019. Microsoft further objects to this Request because it appears designed to invade the attorney-client privilege and attorney work-product protections.

Subject to these objections, Microsoft responds as follows: Microsoft will not search for and produce documents responsive to this Request as they are Privileged Materials.

REQUEST FOR PRODUCTION NO. 69:

Documents sufficient to show all indemnification provisions agreed to by and between **You and OpenAI.**

RESPONSE TO REQUEST FOR PRODUCTION NO. 69:

Microsoft incorporates by reference and reasserts its General Objections to Plaintiffs’ “Definitions” set forth above. Microsoft further objects to this Request as irrelevant and/or vague, ambiguous, overly broad, and not proportional to the needs of the case to the extent it requests information related to the relationship between Microsoft and OpenAI that does not bear any relation to the allegations in the Consolidated Complaint.

Subject to these objections, Microsoft responds as follows: Microsoft has produced the agreements between it and OpenAI related to the Relevant Generative AI Models.

REQUEST FOR PRODUCTION NO. 70:

All Documents Concerning or Relating to whether OpenAI's use of copyrighted material as Training Material complies with copyright law in the United States or elsewhere.

RESPONSE TO REQUEST FOR PRODUCTION NO. 70:

Microsoft incorporates by reference and reasserts its General Objections to Plaintiffs' "Definitions" set forth above. Microsoft objects to this Request on the grounds that it is irrelevant and/or overly broad, unduly burdensome, and not proportional to the needs of the case to the extent that it seeks information concerning use of any copyrighted material, owned by anyone, at any time, and without any subject matter or geographic limitation. Microsoft also objects to the Request as overly broad, unduly burdensome, and not proportional to the needs of the case to the extent it seeks "[a]ll Documents." Microsoft further objects to this Request to the extent it assumes that Microsoft trained or otherwise developed OpenAI's models or applications identified in the Consolidated Complaint. Microsoft did not train the Relevant Generative AI Models, does not possess the training datasets OpenAI used to train the Relevant Generative AI Models, and is unaware of the contents of the training datasets used by OpenAI to train the Relevant Generative AI Models. Microsoft further objects to this Request as irrelevant and/or overly broad and not proportional to the needs of the case to the extent it requests information regarding the copyright laws in any jurisdiction other than the United States. Microsoft further objects to this Request as irrelevant and/or vague, ambiguous, overly broad, and not proportional to the needs of the case because it requests information concerning all of OpenAI's training datasets, pre-training datasets, and/or fine-tuning datasets, regardless of whether they have been identified in the Consolidated Complaint and/or bear any relation to the LLMs at issue in these Actions. Microsoft further objects to this Request because it appears designed to invade the attorney-client privilege and attorney work-product protections.

Subject to these objections, Microsoft responds as follows: Microsoft will not search for and produce documents responsive to this Request as they are Privileged Materials.

REQUEST FOR PRODUCTION NO. 71:

All **Documents You** received from OpenAI in the course of **Your** due diligence regarding each of **Your** investments in **OpenAI**.

RESPONSE TO REQUEST FOR PRODUCTION NO. 71:

Microsoft incorporates by reference and reasserts its General Objections to Plaintiffs' "Definitions" set forth above. Microsoft also objects to the Request as overly broad, unduly burdensome, and not proportional to the needs of the case to the extent it seeks "[a]ll Documents." Microsoft further objects to this Request as irrelevant and/or vague, ambiguous, overly broad, and not proportional to the needs of the case to the extent it requests information related to the relationship between Microsoft and OpenAI that does not bear any relation to the allegations in the Consolidated Complaint.

Subject to these objections, Microsoft responds as follows: Microsoft will conduct a reasonable search and produce responsive, nonprivileged documents, if any, that Microsoft received from OpenAI in the course of Microsoft's due diligence regarding each of Microsoft's investments in OpenAI related to the Relevant Generative AI Models.

REQUEST FOR PRODUCTION NO. 72:

Documents and Communications Concerning or Relating to Your decision to modify **Your** normal due diligence process for the purposes of **Your** investment in **OpenAI**.

RESPONSE TO REQUEST FOR PRODUCTION NO. 72:

Microsoft incorporates by reference and reasserts its General Objections to Plaintiffs' "Definitions" set forth above. Microsoft also objects to the phrase "decision to modify" as vague and ambiguous in that this phrase is undefined. Microsoft further objects to this Request as

irrelevant and/or vague, ambiguous, overly broad, and not proportional to the needs of the case to the extent it requests information related to the relationship between Microsoft and OpenAI that does not bear any relation to the allegations in the Consolidated Complaint.

Subject to these objections, Microsoft responds as follows: Microsoft will not currently search for and produce documents responsive to this Request but is willing to meet and confer.

REQUEST FOR PRODUCTION NO. 73:

Documents sufficient to identify all data that was shared among **You** and **OpenAI** in the course of **Your** partnership with **OpenAI**, including but not limited to data shared pursuant to the terms of the Joint Development and Collaboration Agreement (the “JDCA”) (initial JDCA produced at MSFT_AICPY_000000345) since June 2019.

RESPONSE TO REQUEST FOR PRODUCTION NO. 73:

Microsoft incorporates by reference and reasserts its General Objections to Plaintiffs’ “Definitions” set forth above. Microsoft further objects to this Request as irrelevant and/or vague, ambiguous, overly broad, and not proportional to the needs of the case to the extent it requests information related to the relationship between Microsoft and OpenAI that does not bear any relation to the allegations in the Consolidated Complaint. Microsoft also objects to the Request as vague and ambiguous to the extent it requests “data” that is undefined.

Subject to these objections, Microsoft responds as follows: Microsoft will not currently search for documents responsive to this Request but is willing to meet and confer regarding the scope of this Request.

REQUEST FOR PRODUCTION NO. 74:

Documents sufficient to identify each of **Your** employees who at any point had access to “Restricted Access Model Training Code” pursuant to Section 3(e) of the Second Amended and Restated JDCA (the “Second Amended JDCA”) (MSFT_AICPY_000000575).

RESPONSE TO REQUEST FOR PRODUCTION NO. 74:

Microsoft incorporates by reference and reasserts its General Objections to Plaintiffs’ “Definitions” set forth above. Microsoft further objects to this Request irrelevant and/or as vague, ambiguous, overly broad, and not proportional to the needs of the case to the extent it requests information related to the relationship between Microsoft and OpenAI that does not bear any relation to the allegations in the Consolidated Complaint. Microsoft objects to this Request as vague and ambiguous because the term “access” is undefined.

Subject to these objections, Microsoft responds as follows: Microsoft will conduct a reasonable search and produce responsive, nonprivileged documents, if any, sufficient to show all Microsoft employees who at any point had access to any “Restricted Access Model Training Code” related to the Relevant Generative AI Models.

REQUEST FOR PRODUCTION NO. 75:

Documents sufficient to identify any “secure repository” **You** identified to **OpenAI** pursuant to Section 3(e)(iv) the Amended and Restated JDCA (MSFT_AICPY_000004210).

RESPONSE FOR REQUEST FOR PRODUCTION NO. 75:

Microsoft incorporates by reference and reasserts its General Objections to Plaintiffs’ “Definitions” set forth above. Microsoft further objects to this Request as irrelevant and/or vague, ambiguous, overly broad, and not proportional to the needs of the case to the extent it requests information related to the relationship between Microsoft and OpenAI that does not bear any relation to the allegations in the Consolidated Complaint.

Subject to these objections, Microsoft responds as follows: Microsoft will conduct a reasonable search and produce responsive, nonprivileged documents, if any, sufficient to show any secure repository Microsoft identified to OpenAI related to the Relevant Generative AI Models.

REQUEST FOR PRODUCTION NO. 76:

Documents sufficient to identify all individuals involved in copying the “Development Materials”—as defined in the Second Amended JDCA—for storage in **Your** secure repository, as set forth in Section 3(h)(iv) the Second Amended JDCA.

RESPONSE TO REQUEST FOR PRODUCTION NO. 76:

Microsoft incorporates by reference and reasserts its General Objections to Plaintiffs’ “Definitions” set forth above. Microsoft further objects to this Request as irrelevant and/or vague, ambiguous, overly broad, and not proportional to the needs of the case to the extent it requests information related to the relationship between Microsoft and OpenAI that does not bear any relation to the allegations in the Consolidated Complaint.

Subject to these objections, Microsoft responds as follows: Microsoft will conduct a reasonable search and produce responsive, nonprivileged documents, if any, sufficient to identify any individuals involved in copying any “Development Materials,” as defined in the Second Amended JDCA, for storage in any secure repository of Microsoft related to the Relevant Generative AI Models.

REQUEST FOR PRODUCTION NO. 77:

All **Documents Concerning** or **Relating to** the Data Working Group, as referenced in the Second Amended JDCA, including but not limited to all **Communications** among members of, meeting minutes of, and data policies compiled by the Data Working Group.

RESPONSE FOR REQUEST FOR PRODUCTION NO. 77:

Microsoft incorporates by reference and reasserts its General Objections to Plaintiffs’ “Definitions” set forth above. Microsoft also objects to the Request as overly broad, unduly burdensome, and not proportional to the needs of the case to the extent it seeks “[a]ll Documents” and “all Communications.” Microsoft further objects to this Request as irrelevant and/or vague,

ambiguous, overly broad, and not proportional to the needs of the case to the extent it requests information related to the relationship between Microsoft and OpenAI that does not bear any relation to the allegations in the Consolidated Complaint.

Subject to these objections, Microsoft responds as follows: Microsoft will conduct a reasonable search and produce responsive, nonprivileged documents, if any, concerning Microsoft's involvement in the Data Working Group as it relates to the Relevant Generative AI Models.

REQUEST FOR PRODUCTION NO. 78:

All **Documents** created in the course of the Governing Board's—as referenced in the Second Amended JDCA—duties, including but not limited to meeting minutes, memorandum and **Communications**.

RESPONSE TO REQUEST FOR PRODUCTION NO. 78:

Microsoft incorporates by reference and reasserts its General Objections to Plaintiffs' "Definitions" set forth above. Microsoft also objects to the Request as overly broad, unduly burdensome, and not proportional to the needs of the case to the extent it seeks "[a]ll Documents." Microsoft further objects to this Request as irrelevant and/or vague, ambiguous, overly broad, and not proportional to the needs of the case to the extent it requests information related to the relationship between Microsoft and OpenAI that does not bear any relation to the allegations in the Consolidated Complaint.

Subject to these objections, Microsoft responds as follows: Microsoft will conduct a reasonable search and produce responsive, nonprivileged documents, if any, concerning Microsoft's involvement in Governing Board as it relates to the Relevant Generative AI Models.

Dated: September 13, 2024

Respectfully submitted,

/s/ Jared B. Briant

Jared B. Briant (Admitted Pro Hac Vice)
FAEGRE DRINKER BIDDLE & REATH LLP
1144 Fifteenth Street, Suite 3400
Denver, CO 80202
Telephone: (303) 607-3500
Facsimile: (303) 607-3600
Email: jared.briant@faegredrinker.com

Annette L. Hurst (Admitted Pro Hac Vice)
ORRICK, HERRINGTON & SUTCLIFFE LLP
The Orrick Building, 405 Howard Street
San Francisco, CA 94105-2669
Telephone: (415) 773-5700
Facsimile: (415) 773-5759
Email: ahurst@orrick.com

Sheryl Garko (Admitted Pro Hac Vice)
Laura Najemy (Admitted Pro Hac Vice)
ORRICK, HERRINGTON & SUTCLIFFE LLP
222 Berkley Street
Boston, MA 02116
Telephone: (617) 880-1800
Facsimile: (617) 880-1801
Email: sgarko@orrick.com
lnajemy@orrick.com

Christopher Cariello
Marc Shapiro
ORRICK, HERRINGTON & SUTCLIFFE LLP
51 West 52nd Street
New York, NY 10019
Telephone: (212) 506-3778
Facsimile: (212) 506-5151
Email: ccariello@orrick.com
mrshapiro@orrick.com

Jeffrey S. Jacobson
FAEGRE DRINKER BIDDLE & REATH LLP
1177 Avenue of the Americas, 41st Floor
New York, New York 10036
Telephone: (212) 248-3140
Facsimile: (212) 248-3141
Email: jeffrey.jacobson@faegredrinker.com

Attorneys for Defendant Microsoft Corporation

PROOF OF SERVICE VIA ELECTRONIC MAIL

On September 13, 2024, per the parties' agreements, I directed **DEFENDANT MICROSOFT CORPORATION'S RESPONSES AND OBJECTIONS TO PLAINTIFFS' SECOND SET OF REQUESTS FOR PRODUCTION OF DOCUMENTS** to be served via electronic mail upon:

Alter-AI-TT@simplelists.susmangodfrey.com

Rachel Geman
Anna Josefine Freymann
LIEFF CABRASER HEIMANN &
BERNSTEIN, LLP
250 Hudson Street, 8th Floor
New York, NY 10013-1413
Telephone: 212.355.9500

Scott J. Shoulder
CeCe M. Cole
COWAN DEBAETS ABRAHAMS &
SHEPPARD LLP
60 Broad St., 30th Floor
New York, NY 10004
Telephone: 212.974.7474

Reilly T. Stoler
Ian M. Bensberg
LIEFF CABRASER HEIMANN &
BERNSTEIN LLP
275 Battery Street, 29th Floor
San Francisco, CA 94111-3339
Telephone: 415.956.1000

J. Craig Smyser
Charlotte Lepic
SUSMAN GODFREY L.L.P.
One Manhattan West
Ste. 50th Fl.
New York, NY 10001
Telephone: 212.336.8330

Justin A. Nelson
Alejandra C. Salinas
SUSMAN GODFREY L.L.P.
1000 Louisiana Street, Suite 5100
Houston, TX 77002
Telephone: 713.651.9366

Rohit D. Nath
SUSMAN GODFREY L.L.P.
1900 Avenue of the Stars, Suite 1400
Los Angeles, California 90067
Telephone: 310.789.3100

Wesley Dozier
LIEFF CABRASER HEIMANN &
BERNSTEIN LLP
222 2nd Avenue, Suite 1640
Nashville, TN 37201
Telephone: 615.313.9000

KVPOAI@keker.com

Robert A. Van Nest
R. James Slaughter
Katie L. Joyce
Michelle S. Ybarra
Nicholas S. Goldberg

Thomas E. Gorman
Paven Malhotra
Christopher S. Sun
KEKER & VAN NEST LLP
633 Battery Street
San Francisco, CA 94111
Telephone: 415.391.5400

OpenAICopyrightLitigation.lwteam@lw.com

Sarang Damle
Elana N. Dawson
LATHAM & WATKINS LLP
555 Eleventh Street NW, Suite 1000
Washington, DC 20004
Telephone: 202.637.2200

Andrew Gass
Joseph Richard Wetzel, Jr.
LATHAM & WATKINS, LLP
505 Montgomery Street, Suite 2000
San Francisco, CA 94111
Telephone: 415.391.0600

Allison Levine Stillman
Rachel R. Blitzler
LATHAM & WATKINS LLP
1271 Avenue of the Americas
New York, NY 10020
Telephone: 212.906.1200

OpenAICopyright@mofo.com

Joseph C. Gratz
Vera Ranieri
Andrew L. Perito
Tiffany Cheung
MORRISON & FOERSTER LLP
425 Market Street
San Francisco, CA 94105-2482
Telephone: 415.268.7522

Allyson R. Bennett
Rose S. Lee
MORRISON & FOERSTER LLP
707 Wilshire Boulevard
Los Angeles, CA 90017-3543
Telephone: 213.892.5200

Eric Nikolaides
MORRISON & FOERSTER LLP
250 W. 55th Street
New York, NY 10019
Telephone: 212.468.8000

/s/ Shelley M. Meyer
Shelley M. Meyer